# Denormalization RDBMS

## Goal

Understand the interest of the denormalization process in the context of a relational database. We will be using the PostgreSQL system

### Download

Retrieve the use case on the course's e-learning Web page (denorm_td1.zip)

### 1. Modeling

Analyze the three files of the zip archive et propose an Entity Relationship (ER) diagram (without any optimization). Create the relational model from your ER diagram and translate it into SQL DDL instructions.

### 2. Integration

Once the tables have been created, perform some transformations on the data files in order to integrate the data into your tables. You can perform these operations with the programming (script) languages of your choice (Python Java, Gawk).

### 3. Querying

Write a SQL query that retrieves the cip7 and cis ids, the associated drug name and the pharmaceutical laboratory name.
Execute the query with the EXPLAIN ANALYZE mode.
Then, adapt the query to access the drug with the 69309629 ciis value. Execute with an EXPLAIN ANALYZE.

### 4. Indexing

Create some indexes to improve the execution of the previous query. Verify that your approach is more efficient with an EXPLAIN ANALYZE execution.

### 5. Denormalization

In our application, the previous query is very important and we thus need to guarantee « optimum » performance. Propose a denormalization such that your query can be written without any joins. Integrate your data according to your denormalization approach. Do not heistate to create additional tables. Modify the two previous queries and compare performances.
Propose another solution to improve the performance of a query retrieving data of the 69309629 cis value.

### 6. Modification

Considering the impact of your model when one is updating (via insert, update, delete SQL queries) a drug of the cis table.

---

### Appendix

The elements in the CIS file are :
- the cis id (an identifying code for a drug speciality)
- the name of the drug
- the pharmaceutical form of the drug
- administration routes ( with a « ; »separator between each entry)
- the administrative status regarding the authorization to sell that drug
- procedure type
- commercialization status
- authorization date
- a status with different possible values : « Alerte » or « warning disponibilité »
- European identification value (values can be separated by « ; » symbols)
- a boolean value

Considering the CIS_CIP File, the fields are :
- cis id
- cip7 id
- Presentation label
- Admin status of the presentation
- Commercialization status
- Commercialization date
- cip13 id
- Collectivity aggreement ("oui", "non" ou « inconnu »)
- Reimbursement rate
- price
- Official text regarding reimbursement

In the CIS_COMPO file, we get the following information :
- cis id
- a label
- substance id
- substance name
- dosage reference
- component nature (principe actif : « SA » ou fraction thérapeutique : « ST »)
- SA or FT number